PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-143486
(43)Date of publication of application : 28.05.1999

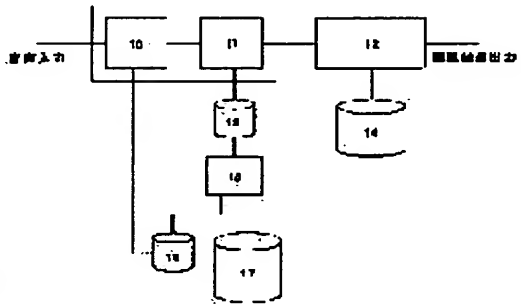
(51)Int.CI. G10L 3/00
G10L 3/00
G10L 3/00

(21)Application number : 09-306887
(22)Date of filing : 10.11.1997

(71)Applicant : FUJI XEROX CO LTD
(72)Inventor : SUMIYA KAZUHIKO
SAITO NOBUYUKI

(54) DEVICE AND METHOD ADAPTABLE FOR SPEAKER

(57)Abstract:
PROBLEM TO BE SOLVED: To realize the speaker-adaptable capacity excellent in accuracy in the speaker adaptation using the maximum posterior probability estimating method.
SOLUTION: A set 17 of a large number of speaker models independent from an adaptable speaker is prepared using an acoustic analyzing means equivalent to an acoustic analysis part 10. Then, the sound of the adaptable speaker is inputted, and analyzed by the acoustic analysis part 10, the distribution of the feature parameter vector of the adaptable speaker is obtained, and preserved as the sample data 16 for adaptation. An datable model preparation part 15 measures the distance between the adaptable speaker model preserved as the sample data 16 for adaptation and a large number (N-pieces) of speaker models preserved as the set 17 of the speaker model, and the speaker models of M pieces are selected in the order of smaller distance to the adaptable speaker model. The weighted addition of the selected speaker models of M pieces is achieved to determine the initial model.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

This Page Blank (uspto)

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

This Page Blank (uspto)

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平11－143486

(43)公開日　平成11年(1999)５月28日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ１０Ｌ　3/00	５３５	Ｇ１０Ｌ　3/00	５３５
	５２１		５２１Ｆ
	５３１		５３１Ｆ
			５３１Ｋ

審査請求　未請求　請求項の数４　ＯＬ　（全 10 頁）

(21)出願番号　特願平9－306887

(22)出願日　平成９年(1997)11月10日

(71)出願人　000005496
富士ゼロックス株式会社
東京都港区赤坂二丁目17番22号

(72)発明者　住谷　和彦
神奈川県足柄上郡中井町境430　グリーンテクなかい　富士ゼロックス株式会社内

(72)発明者　斎藤　伸行
神奈川県足柄上郡中井町境430　グリーンテクなかい　富士ゼロックス株式会社内

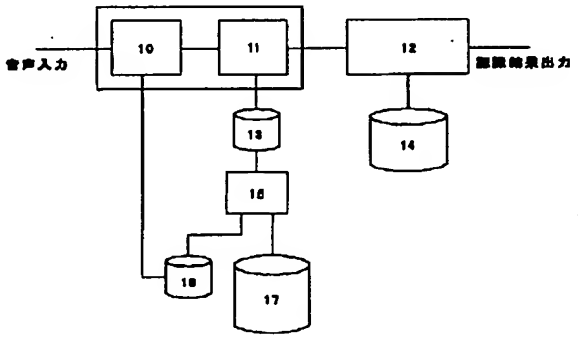
(74)代理人　弁理士　澤田　俊夫

(54)【発明の名称】　話者適応装置および方法

(57)【要約】

【課題】　最大事後確率推定法を用いた話者適応において、精度の高い話者適応の能力を実現する。

【解決手段】　音響解析部１０と同等の音響解析手段を使って、適応対象話者と独立した多数の話者モデルの集合１７を用意する。次に、適応対象の話者の音声を入力し、音響解析部１０で分析して適応対象話者の特徴パラメータベクトルの分布を求め、適応用サンプル・データ１６として保存する。適応モデル作成部１５では、適応用サンプル・データ１６として保存されている適応対象の話者モデルと話者モデルの集合１７として保存されている多数（Ｎ個）の話者モデルとの間の距離を測定し、適応対象の話者モデルとの距離が近い順にＭ個の話者モデルを選ぶ。そして、選択したＭ個の話者モデルの重み付き加算値を行い初期のモデルを決定する。



1

【特許請求の範囲】

【請求項１】　初期話者モデルと適応学習用データとを用いて、最大事後確率推定法によって話者モデルのパラメータを再推定し、話者適応を行う話者適応装置において、事前に、多数の話者から多数の初期話者モデルを作成し、適応対象の話者が発声した適応用学習データから、その適応対象話者の特徴を抽出し、前記多数の初期話者モデルの中から、前記適応対象話者の特徴に距離的に最も近い方からＮ個の話者モデルを選択し、選択されたＮ個の話者モデルの各々を事前に仮定された分布として、適応用学習データを使って話者モデルのパラメータを推定し、その推定されたパラメータを持つＮ個の話者モデルを混合加算することにより適応対象話者の音声モデルを作成することを特徴とする話者適応装置。

【請求項２】　混合するＮ個の話者モデルは、適応対象話者の特徴ベクトルとの距離に応じて、重み付けされることを特徴とする請求項１記載の話者適応装置。

【請求項３】　適応対象の話者と距離的に最も近い話者モデルとの距離を基底距離とするとき、適応対象の話者との距離が前記基底距離と比較して一定範囲以内である話者モデルを選択することにより、混合する話者モデルの個数Ｎを可変とすることを特徴とする請求項１または２記載の話者適応装置。

【請求項４】　初期話者モデルと適応学習用データとを用いて、最大事後確率推定法によって話者モデルのパラメータを再推定し、話者適応を行う話者適応方法において、事前に、多数の話者から多数の初期話者モデルを作成し、適応対象の話者が発声した適応用学習データから、その適応対象話者の特徴を抽出し、前記多数の初期話者モデルの中から、前記適応対象話者の特徴に距離的に最も近い方からＮ個の話者モデルを選択し、選択されたＮ個の話者モデルの各々を事前に仮定された分布として、適応用学習データを使って話者モデルのパラメータを推定し、その推定されたパラメータを持つＮ個の話者モデルを混合加算することにより適応対象話者の音声モデルを作成することを特徴とする話者適応方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、適応対象話者が発声した音声を学習用データとして、初期の音声モデルを修正し、話者に適応させた音声モデルを作成する、話者適応技術に関する。

【０００２】

【従来技術】隠れマルコフモデル（Ｈｉｄｄｅｎ　Ｍａｒｋｏｖ　Ｍｏｄｅｌ、以降HMMと略する）は、音声のスペクトル的、時間的な変動に対処しやすく、高い認識精度を実現できることから、音声認識において広く用いられている。HMMは、状態間の遷移確率、状態遷移に伴うシンボルの出力確率を持った状態遷移モデルであり、音声信号のような時間とともに連続的に変化する信

2

号をモデル化するには、左から右に状態が遷移する、所謂ｌｅｆｔ－ｔｏ－ｒｉｇｈｔ型モデルが適当である。図１に状態数４の場合のｌｅｆｔ－ｔｏ－ｒｉｇｈｔ型のHMMの例を示す。ここで、１，２，３，４は状態を表し、a_{ij}（ｉ，ｊ＝１，…４）は、状態ｉから状態ｊに遷移する確率を示している。また、b_i（ｏ）は、状態遷移に伴って状態ｊにおいてシンボルｏが観測される確率を示している。音声認識の場合、このシンボルとしては、通常、特徴パラメータ・ベクトルが使われる。

【０００３】HMMを使った音声認識では、認識の対象となる音声の単位（例えば、音素、音韻、音節、単語）ごとに、HMMによる音声モデルを用意し、事前に訓練用の音声データを用いて、状態遷移確率a_{ij}、特徴パラメータベクトルの出力確率b_i（ｏ）といったモデル・パラメータを決定しておく。そして、認識時には、入力された音声を分析し、特徴パラメータベクトルの系列に変換し、そのパラメータベクトルの系列を観測する可能性が最も高くなるモデルに対応する音声単位の列を決定し、それを認識結果とする。

【０００４】HMMのような統計モデルでは、一般に、学習用のデータを増やすことにより、モデルの精度を高めることができる。そこで、対象となる話者が発声した大量のデータを用いることによって、その話者のための精度の高いモデルを作成することができる。しかし、そのためには、使用に先立って大量の音声データを収集する必要があり、そのために話者に大量の発声を要求することになり、実用上の障害になっていた。

【０００５】一方、事前に不特定多数の話者の音声データを収集し、それに基づいて、話者に依存しない標準的な話者モデルを作ることが考えられる。この不特定話者を対象としたモデルを用いれば、使用者が使用に先立って話者モデルを訓練するために大量の発声を行う必要がなく、すぐに使用を開始できるという利点がある。しかし、話者ごとに適合させたモデルではないので、認識精度が十分ではない。一般に、音声認識では、個人に即した音声モデルを用意することにより、認識の精度を高めることができるが、その音声モデルを作成するためには多数の発話を事前に用意する必要があり、認識精度と手間がトレードオフの関係になっている。

【０００６】こうしたことから、少量の話者固有の発声データだけで、使用を開始することができ、さらに話者固有の発声データを追加することによって逐次的に認識精度を向上させる事を狙った「話者適応」が、音声認識を応用するある局面では注目されている。例えばディクテーションのような、ある程度個人的に使用されるものでは、簡単な手続きで使用することができ、また、使用につれて、逐次、認識の精度が向上していくことが望ましく、この話者適応が有効であろう。

【０００７】「話者適応」の手法の一つとして、予め求めておいた初期の音声モデルを、実際の話者や使用環境

の特徴を取り込んで、修正することによって実現する手法があり、その中でも、ベイズ推定に基づいた音声モデルのパラメータの再推定が試みられ、効果をあげている。ベイズ推定は、ベイズの定理に基づいて、パラメータを決定するものであり、以下の考え方による。

【０００８】ある事象Ａという結果を生じさせた原因H_iの可能性Ｐ（H_i｜Ａ）は、ベイズの定理により、原因の確率Ｐ（H_i）に、その原因からある事象が発生する確率Ｐ（Ａ｜H_i）をかけたものによって計算される。この時、Ｐ（H_i）は事前確率とよばれ、Ｐ（H_i｜Ａ）は事後確率と呼ばれる。つまり、事前に予測した事前確率Ｐ（H_i）とサンプルの確率Ｐ（Ａ｜H_i）によって事後確率が決定すると考える。

【０００９】原因として分布のパラメータをとり、結果としてサンプルされるデータをとる。そうすると、分布のパラメータの事後確率は、分布のパラメータの予測値である事前確率とサンプルデータから得られる結果の確率から得られることになる。そこで、モデルのパラメータについての正しい予測があれば、事後確率の決定にそれを効果的に取り込むことができる。

【００１０】一般に、結果の確率に対して、事前確率と事後確率が同一の分布族に属していれば、サンプルは分布族内の変換を起こすだけであり、数学的な取り扱いが容易になる。このときの事前確率と事後確率の分布族は、自然な共役分布と呼ばれるが、ガウス分布Ｎ（θ，＊σ^2）の平均のパラメータθについて、ガウス分布Ｎ（λ，τ^2）は自然な共役分布であることが知られている。

【００１１】ベイズの定理に基づく枠組みの中で、パラメータを推定し話者適応を行う方法が、最大事後確率推定法として、例えば、Ｃｈｉｎ－Ｈｕｉ Ｌｅｅ，Ｃｈｉｈ－Ｈｅｎｇ Ｌｉｎ ａｎｄ Ｂｉｉｎｇ－Ｈｗａｎｇ Ｊｕａｎｇ，”Ａ Ｓｔｕｄｙ ｏｎ Ｓｐｅａｋｅｒ Ａｄａｐｔａｔｉｏｎ ｏｆ ｔｈｅ Ｐａｒａｍｅｔｅｒｓ ｏｆ Ｃｏｎｔｉｎｕｏｕｓ Ｄｅｎｓｉｔｙ Ｈｉｄｄｅｎ Ｍａｒｋｏｖ Ｍｏｄｅｌｓ”，ＩＥＥＥ Ｔｒａｎｓａｃｔｉｏｎｓ ｏｎ Ｓｉｇｎａｌ Ｐｒｏｃｅｓｓｉｎｇ，Ｖｏｌ．３９，Ｎｏ．４，Ａｐｒｉｌ，１９９１（以後文献１とする）で、開示されている。これによれば、平均値パラメータμが事前確率Ｐ。（μ）に従い、その分散σ^2が既知で固定とすると、μの共役な事前分布は、平均値ν、分散τ^2を持ったガウス分布であり、これを使うと、パラメータの最大事後確率推定値は、

【００１２】

【数１】

$$\mu_{MAP}=\frac{n\tau^2}{\sigma^2+n\tau^2}\bar{o}+\frac{\sigma^2}{\sigma^2+n\tau^2}\nu \qquad (1)$$

（文章中の表記の制限により「$\bar{o}$」を「o」で表すこともある。）

で与えられる。ここで、ｎは対応するＨＭＭの状態において観測される訓練用サンプルの個数、oはサンプルの平均である。つまり、平均値の最大事後確率推定法による推定値は、事前分布の平均値νとサンプルの平均oの重み付き平均で与えられる。（１）から明らかに、ｎが０のとき、つまりサンプルが全くない場合には、推定値は事前分布の平均値νのままである。また、ｎが十分大きいとき、つまり、サンプル数が十分多いときは、推定値はサンプルの平均oに近づく。このように、最大事後確率推定法では、事前に予測した分布を取り込んで、サンプルデータの個数に応じてモデルのパラメータを訓練用サンプルの特性に漸近的に近づけることができるので、使用に応じて逐次話者モデルの精度向上を図ることができ、理想的な話者適応を実現できる。

【００１３】文献１では、ベイズ推定による他のパラメータに対する話者適応例として、分散の場合と平均と分散双方の場合についても述べられているが、平均値の話者適応が認識性能に対して効果が高いことが示されている。

【００１４】

【発明が解決しようとする課題】この最大事後確率推定法によるパラメータ推定を使った話者適応方式が、これまでに開示されている。特開平８－９５５９２号公報で開示されている従来例１では、標準的音素モデルを用いて、ある特定の話者に音素モデルを合わせ込む話者適応を行っているが、標準的な初期音声モデルとしては、老若男女いろいろな話者が発声した音声データを用いて予め学習しておいた、不特定多数の話者の音声を認識対象とした不特定話者モデルを用いるとしている。この従来例１で示している不特定話者モデルを用いた適応方式では、事前分布として用いる標準的な初期音声モデルの分布が広がっており、少量の学習データで、それを十分に補償し、適応対象の話者に適応するようパラメータを補正することは困難である。

【００１５】また、特開平８－１１０７９２号公報で開示されている従来例２では、木構造クラスタリングモデルにより作成した話者クラスタを用いて初期話者モデルを作成するとしている。この方法では、初期話者モデルの分布をある程度絞る事ができるが、クラスタの中心が適応対象となる話者モデルのベクトルの中から、大きくずれている場合には、補正の効果が小さく、十分な適応ができない恐れがある。

【００１６】最大事後確率推定法は、前述したように、話者の発声サンプルデータを用いて、初期話者モデルのパラメータを修正し、話者に適応させていくもので、サンプルデータの数に応じて漸次モデルの精度を向上させることができるが、この初期の話者モデルをどう選ぶかが、適応の能力に大きく影響する。本発明は、こうした

点に鑑みなされたもので、最大事後確率推定法を用いた話者適応において、精度の高い話者適応の能力を実現するための手段を与えるものである。本発明は、最適な初期の話者モデルを選択することにより、適応対象の特定の話者からの少量の発声データを用いて、その話者に対する精度の高い音声モデルを作成する話者適応装置を実現することを目的とするものであり、その話者適応装置を用いることにより、精度の高い話者適応の能力を持った音声認識装置を実現することができる。

【００１７】

【課題を解決するための手段】本発明は、初期話者モデルと適応学習用データを用いて、最大事後確率推定法によって話者モデルのパラメータを再推定し、話者適応を行う装置において、多数の話者から作成した多数（Ｎ個）の話者モデルを事前に用意し、それらの多数の話者モデルの中から、適応対象の話者に距離的に近いＭ個の話者モデルを選択し、そうして選択されたＭ個の話者モデルを混合して初期の話者モデルを構成することを特徴とする。また、そのとき、個数Ｍ（Ｍ＜＜Ｎ）を適応対象話者と事前に用意されたＮ個の個々の話者モデルとの距離の関係に基づいて決定する。これにより、分布に応じた精度の高い初期モデルを決定することができるとともに、不必要なパラメータの増加を押さえることができる。

【００１８】この構成においては、混合するＮ個の話者モデルは、適応対象話者の特徴ベクトルとの距離に応じて、重み付けされるようにしてもよい。

【００１９】また、適応対象の話者と距離的に最も近い話者モデルとの距離を基底距離とするとき、適応対象の話者との距離が前記基底距離と比較して一定範囲以内である話者モデルを選択することにより、混合する話者モデルの個数Ｎを可変とするようにしてもよい。

【００２０】また、本発明は方法としても実現でき、また少なくともその一部をコンピュータプログラム製品としても実現できる。

【００２１】

【発明の実施の態様】以下、本発明の実施例について説明する。

【００２２】図２は、本発明による話者適応装置とそれを用いた音声認識システムの実施例をブロック図で示したものである。音声認識システムは、入力された音声を音響解析部１０で分析し、特徴パラメータ・ベクトルの系列を抽出する。抽出した特徴パラメータ・ベクトルの系列を、音韻照合部１１において、言語モデル１４からの粗い情報を参照しながら、音声モデル１３と照合し、複数の音韻系列の候補を作成する。こうしてできた複数の音韻系列の候補を、言語認識部１２で言語モデル１４からの細かい情報を使って再評価し、最終的な認識結果を確定する。上記の話者に適応した音声モデル１３を作る手段が本発明による話者適応化装置であり、その方法と構成を以下に述べる。

【００２３】この実施例では、まず、事前に多数の話者の音声を収集し、図示していない、音響解析部１０と同等の音響解析手段を使って、適応対象話者と独立した多数の話者モデルの集合１７を用意しておく。つまり、Ｎ人の話者の音声を収集することにより、Ｎ個の分布を予め計算して用意しておく。図３は、これらの多数の話者モデルの特徴パラメータベクトルの出力確率分布を模式的に破線で示したものである。説明の都合上、特徴パラメータベクトルを２次元として表示しているが、この特徴パラメータベクトルは、実際には３０次元程度の多次元ベクトルとするのが普通である。各話者モデルは、図３に示すように、相互に重なり合いながら、多次元空間内に分布している。次に、話者適応のために、適応対象の話者の音声を入力し、音響解析部１０で分析して適応対象話者の特徴パラメータベクトルの分布を求め、適応用サンプル・データ１６として保存する。図４は、この適応用サンプル・データから得られた適応対象話者の特徴パラメータベクトルの出力確率の分布の様子の例を実線で示したもので、先に求めた多数の特定話者モデルの分布との関係を示したものである。以上のようにして適応対象の話者モデルと多数（Ｎ個）の話者モデルとを決定しておき、適応モデル作成部１５において、適応化した音声モデル１３を作成する。それは次の手順で行う。適応モデル作成部１５では、まず適応用サンプル・データ１６として保存されている適応対象の話者モデルと話者モデルの集合１７として保存されている多数（Ｎ個）の話者モデルとの間の距離を測定する。話者モデルとして、多変量のガウス分布を仮定すると、この距離としては、適応対象の話者モデルの中心ベクトルと特定話者モデルとのマハラノビス距離を用いることができる。この距離を使って、適応対象の話者モデルとの距離が近い順にＭ個の話者モデルを選ぶ。図４は、Ｍを３とした場合について、適応対象の話者モデルの最近傍にある、ｎ番目、ｎ＋１番目とｎ＋２番目のモデルが選ばれる様子を説明している。こうして、Ｍ個の話者モデルが選択できると、それらの重み付き加算値として初期のモデルを決定する。

【００２４】なお、この時の重みをこれらの距離に基づいて決定することができる。

【００２５】また、混合するモデルの個数Ｍを固定値とせず、分布に応じて変化させるようにしてもよい。初期モデルの候補として選択するモデルが不確かな場合は、混合するモデルの個数Ｍを多くし、確かな場合は混合するモデルの個数Ｍを少なくする。そのために、適応対象の話者モデルの中心ベクトルとＮ個の特定話者モデルとのマハラノビス距離を測定し、その距離が、全モデルに対する距離の中で最も短いものと比べてある一定範囲内であるモデルを選ぶ。そうして選ばれたモデルの個数をＭ　個であるとき、先の例と同様にしてＭ　個のモデル

の重み付き加算による混合で初期のモデルを作成する。このときも、先の例と同様に、重みはこれらの距離に基づいて決定することができる。こうすることによって、曖昧性が強い、不確かな初期モデルの候補に関しては、混合数を多くして細かいモデルを作成し、逆に、曖昧性が低く、確度の高い初期モデルの候補に関しては、混合数を少なくして計算量を削減したモデルを作ることができる。上記の一定範囲内としては、適応対象の話者モデルの中心ベクトルと全モデルに対する距離の中で最も短いものをD_{min}とするとき、その距離とD_{min}の比が一定値以下のもの、或いは、その距離とD_{min}の差が一定値以下のものを選べばよい。

【００２６】以下に具体例について詳しく述べる。

【００２７】［具体例１］ここでは、連続音声認識に対応した話者適応方式を例として示す。連続音声認識のためには、認識の単位となる音声の単位を音素レベルとするのが適当である。そこで、ここでは、音素毎にＨＭＭを作成する。ＨＭＭの作成は、音声データを使ってＨＭＭのパラメータを決定する訓練と呼ばれる手続きを実行することにより行われる。連続的に発声された発声データを用いて音素レベルのＨＭＭを訓練するには、発声された音素列に関して、先行する音素のＨＭＭの最終状態を後続する音素のＨＭＭの初期状態につなげて訓練を行う。

【００２８】本発明を実施するためには、事前に多数の話者からの発声データを収集し、認識単位となる音声単位毎にＨＭＭを作成する必要がある。音声信号のような時間とともに連続的に変化する信号をモデル化するに＊は、前述したように、左から右に状態が遷移する、所謂ｌｅｆｔ－ｔｏ－ｒｉｇｈｔ型のＨＭＭが適当であるので、図１のような、例えば４状態のＨＭＭを使用する。図１において、１，２，３，４は状態を表し、a_{ij}（ｉ，ｊ＝１，…４）は、状態ｉから状態ｊに遷移する確率を示している。状態ｊにおいて、モデルから観測される事象をb_i（ｏ）であらわす。この観測事象の対象である音声の特徴パラメータベクトルは、連続信号であるので、b_i（ｏ）は以下のような連続確率密度関数で表すことができる。

【００２９】

【数２】

$$b_j(o) = N[o, \mu_j, \Sigma_j] \qquad (2)$$

ここで、ｏは観測ベクトル、Ｎ［…］は、平均値ベクトルμ_i、分散・共分散行列Σ_iのガウス分布関数である。

【００３０】個々のモデルにおけるパラメータ、a_{ij}（ｉ，ｊ＝１，…４）やb_i（ｏ）（ｊ＝１，…４）は、公知のＢａｕｍ－Ｗｅｌｃｈ法を使った繰り返し計算により求めることができる。このＢａｕｍ－Ｗｅｌｃｈ法による再推定については、例えば、文献、Ｌａｗｒｅｎｃｅ Ｒａｂｉｎｅｒ，Ｂｉｉｎｇ－Ｈｗａｎｇ Ｊｕａｎｇ，"Ｆｕｎｄａｍｅｎｔａｌｓ ｏｆ Ｓｐｅｅｃｈ Ｒｅｃｏｇｎｉｔｉｏｎ"，Ｐｒｅｎｔｉｃｅ Ｈａｌｌ ＰＴＲ９３で詳しく述べられている。その概要を示すと以下の通りである。Ｂａｕｍ－Ｗｅｌｃｈ法では、

【００３１】

【数３】

$$a_{ij} = \frac{\text{状態 } i \text{から状態 } j \text{に遷移する回数の期待値}}{\text{状態 } i \text{から遷移する回数の期待値}} \qquad (3)$$

$$b_j(o) = \frac{\text{状態 } j \text{でベクトル } o \text{を観測する回数の期待値}}{\text{状態 } j \text{にある回数の期待値}} \qquad (4)$$

として、パラメータを再推定し、推定されたパラメータを入れ替えて推定の計算を繰り返しすることにより、ｏが観測される確率が極大となるようなモデルを作成することができる。

【００３２】このとき、b_i（ｏ）として、式（２）の※ようなガウス分布を仮定すると、μ_i、Σ_iに対する再推定の式は、

【００３３】

【数４】

$$\mu_j = \frac{\sum_{t=1}^{T} \gamma_t(j) \cdot o_t}{\sum_{t=1}^{T} \gamma_t(j)} \qquad (5)$$

$$\Sigma_j = \frac{\sum_{t=1}^{T} \gamma_t(j) \cdot (o - \mu_j)(o - \mu_j)^t}{\sum_{t=1}^{T} \gamma_t(j)} \qquad (6)$$

となる。ここで、$\gamma_t(j)$は、観測系列oとモデルが与えられたとき、時刻tにおいて状態jに存在する確率である。以上のようにして、事前に作成しておく、複数の話者のモデルが決定する。

【００３４】適応は、以下のステップで行う。まず、適応対象話者から適応学習用の音声データを収集する。そして、これらの適応用学習データに含まれる各音素のＨＭＭについて、上述した方法に従い、パラメータを決定する。そのとき、ビタビ・セグメンテーションにより、音素のＨＭＭの各状態に対応するフレームが決まっているので、それによって、適応学習用データのサンプル数が求められる。

＊【００３５】次に、学習データから得られた、適応対象話者のモデルと、事前に得られている多数の話者モデルとの距離を計算する。この距離としては、平均値ベクトルと話者モデルとのマハラノビス距離を使う。

【００３６】あるｋ番目の話者モデルのｊ番目の状態の分布の平均値ベクトルがμ_{kj}、分散・共分散行列がΣ_{kj}であるとし、適応対象話者のｊ番目の状態の平均値ベクトルμ_jとこのｋ番目のモデルのマハラノビス距離をD_{kj}とするとき、D_{kj}の二乗は

【００３７】

【数５】

$$D_{kj}^2 = (\mu_{kj} - \mu_j)\Sigma_{kj}^{-1}(\mu_{kj} - \mu_j)^t \qquad (7)$$

で求められる。ここで、Σ_{kj}^{-1}は、分散・共分散行列Σ_{kj}の逆行列、$(\mu_{kj}-\mu_j)^t$はベクトル$(\mu_{kj}-\mu_j)$の転置ベクトルである。

【００３８】すべての話者モデルに対してマハラノビス距離D_{kj}を計算し、距離の近い方からM個の話者のモデルを選択する。図４は、これを説明するために、Mが３のときの様子を示したもので、距離の近い方から３つのモデル、ｎ、ｎ＋１、ｎ＋２が選ばれている。但し、状態に関する添字は省略している。

【００３９】さて、この選択されたM個の話者モデルの中の、ｍ（１≦ｍ≦Ｍ）番目の話者モデルの特徴ベクト※ルを考える。そしてその特徴ベクトルのある要素の平均をμ_{mj}、その分散をτ_{mj}^2とする。　今、特徴ベクトルの要素の平均μ_jが、事前分布Ｐ。（μ）を持ち、その分散σ_j^2が既知の固定値であるとする。そうすると、μ_jに対する共役な事前分布は、ガウス分布となる。そこで、ｍ（１≦ｍ≦Ｍ）番目の話者モデルを平均に対する共役事前分布とすると、最大事後確率推定法により、平均の推定値μ_{mj}^{MAP}は、

【００４０】

【数６】

$$\mu_{mj}^{MAP} = \frac{n\tau_{mj}^2}{\sigma_j^2 + n\tau_{mj}^2} o_m + \frac{\sigma_j^2}{\sigma_j^2 + n\tau_{mj}^2}\mu_{mj} \qquad (8)$$

で与えられる。ここで、ｎはサンプルの個数である。

【００４１】つまり、ｍ（１≦ｍ≦Ｍ）番目の話者モデルを初期モデルとした、特徴ベクトルのある要素の平均の推定値はμ_{mj}^{MAP}、事前分布の平均値μ_{mj}と、適応学習において観測されたサンプルの平均o_mの重み付き平均となる。

【００４２】そこで、ｍ番目の話者モデルに対して得られた推定値を用いて、適応対象話者に対するモデルは、次の形で与えられる。

【００４３】

【数７】

$$b_j(o) = \sum_{m=1}^{M} c_{mj}\, N[o, \mu_{mj}^{MAP}, \Sigma_{mj}] \qquad (9)$$

ここで、c_{mj}は、話者モデルのｊ番目の状態における、ｍ番目の話者に対する重みの係数で、これは、前記の過程の中で計算されている距離に基づいて決定することができる。つまり、

【００４４】

【数８】

$$c_{mj} = \frac{\frac{1}{D_{mj}}}{\sum_{m=1}^{M}\frac{1}{D_{mj}}} \qquad (10)$$

とすればよい。このとき、あきらかに

【００４５】

【数９】

$$\sum_{m=1}^{M} c_{mj} = 1 \qquad (11)$$

であり、（９）のは統計的制約を満足するように決定されている。

【００４６】［具体例２］具体例１では、モデルの混合数Mを固定したが、このMの値をモデルの分布に応じて変化させることにより、計算コストに対して適応能力が高い、効率的なモデルを作成する事ができる。つまり、適応対象の話者モデルの中心ベクトルと事前に用意されている多数の特定話者モデルとの距離の分布が、状態ごとに違い、一様ではないときには、その確からしさに応じて、Mの値を変化させるのである。例えば、適応対象の話者モデルの中心ベクトルと事前に用意されている多

数の話者モデルとの距離が、それぞれの話者モデル間で大きく異なり、少数の話者モデルのみに近いときには、それらのモデルのみを混合する分布として利用すれば十分であり、そうすることにより、無駄なパラメータの増加を防ぐことができるからである。

【００４７】具体例２では、適応対象話者のモデルと事前に得られている多数の話者モデルとのマハラノビス距離D_kをとり、それに基づいて混合数を決定する。本発明を実施するためには、まず、具体例１と同様にして、事前に多数の話者からの発声データを収集し、認識単位となる音声単位毎に多数の話者のＨＭＭを作成する。次に、適応対象話者から適応用学習用の音声データを収集し、適応用学習データに含まれる各音素のＨＭＭについて、モデルのパラメータを決定する。そうして、こうして得られた適応対象話者のモデルと、多数の話者モデルとの距離を計算する。距離としては、マハラノビス距離を使う。この具体例２では、適応対象話者のモデルと、多数の話者モデルとの距離のうち最少のものを

【００４８】

【数１０】

$$D_{min} = \min_{1 \le m \le M} D_m \qquad (12)$$

とするとき、D_{min}と比較して規定の範囲の距離D_kをもつモデルｋのみを混合する要素として選択する。

【００４９】この様子を図で説明すると以下のようになる。例えば、図５のように、モデルｎに対する距離D_nが最少であり、モデルｎ＋１、モデルｎ＋２に対する距離D_{n+1}、D_{n+2}がその最少距離と比較してある範囲以内であれば、モデルｎ、モデルｎ＋１、モデルｎ＋２の３つのモデルが、混合するモデルとして選ばれる。また、例えば、図６のように、モデルｎに対する距離が最少であり、モデルｎ＋１に対する距離はと比較してある範囲以内あるが、その次に近いモデルｎ＋２に対する距離が＊と比較してある範囲以内になければ、モデルｎとモデルｎ＋１の２つのモデルのみが混合するモデルとして選択される。

【００５０】この最少距離と比較してある範囲を決定する方法としては、モデルｋとの距離D_kと最少距離D_{min}との比が、一定値δ以下である、つまり

【００５１】

【数１１】

$$\frac{D_k}{D_{min}} < \delta \qquad (13)$$

なるｋ番目のモデルを選択するのが一つの方法である。

【００５２】また、モデルｋとの距離D_kと最少距離D_{min}との差が一定値δ’以下である、つまり

【００５３】

【数１２】

$$\left| D_k - D_{min} \right| < \delta' \qquad (14)$$

なるｋ番目のモデルを選択する事もできる。

【００５４】いずれかの方法で、選択されたモデルの個数をＭ’とするとき、それらの選択された話者モデルに対して得られた推定値を用いて、適応対象話者に対するモデルは、具体例１の場合と同様にして次の形で与えられる。

【００５５】

【数１３】

$$b_j(o) = \sum_{m=1}^{M'} c_{mj}\, N[o, \mu_{mj}^{MAP}, \Sigma_{mj}] \qquad (15)$$

ここで、c_{mj}は、話者モデルのｊ番目の状態における、番目の話者に対する重みの係数で、これは、前記の過程の中で計算されている距離に基づいて決定することができる。つまり、

【００５６】

【数１４】

$$c_{mj} = \frac{\frac{1}{D_{mj}}}{\sum_{m=1}^{M'} \frac{1}{D_{mj}}} \qquad (16)$$

とすればよい。

【００５７】

【発明の効果】本発明は、事前に仮定した初期話者モデルと適応学習用データを用いて、最大事後確率推定法によって話者モデルのパラメータを再推定し、話者適応を行う装置において、初期話者モデルとして適応対象の話者の特性にできるだけ近いと考えられる予測分布を仮定しようとするものである。初期話者モデルに適応対象の話者の特性にできるだけ近い予測分布を用いることにより、少量の適応学習用データで精度の高いモデルのパラメータ推定が行われ、良好な話者適応が実現する。本発明では、事前に得られている多数の特定話者モデルと適応対象話者の距離を測定し、距離的に近いＮ個のモデルを選択的に用いることによりこの作用を効果的に発現させる方法を開示しており、これにより高い適応性を実現する事ができる。

【００５８】また、本発明では、距離的に近いＭ個のモデルを選択するときに、モデルとの距離の相対的な関係により、選択するモデルの個数Ｍを変化させる方法を開示している。これは、仮定する予測分布の分布に応じて最適な混合モデルを選択するもので、これにより、最良の適応能力が発揮されるとともに、計算処理量も減少す

るという効果もある。

【図面の簡単な説明】

【図１】　HMMの例を示す模式図である。

【図２】　本発明の実施例を示すブロック図である。

【図３】　多数の特定話者モデルの分布の例を示す図である。

【図４】　学習データから得られる適応対象話者の分布の例を示す図である。

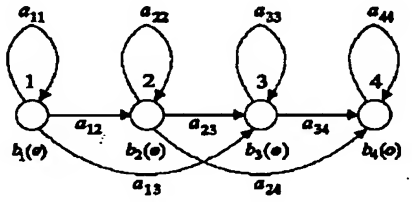
【図５】　適応対象話者モデルの最近傍にある特定話者モデルの分布の例を示す図である。

【図６】　適応対象話者モデルの最近傍にある特定話者モデルの分布の他の例を示す図である。

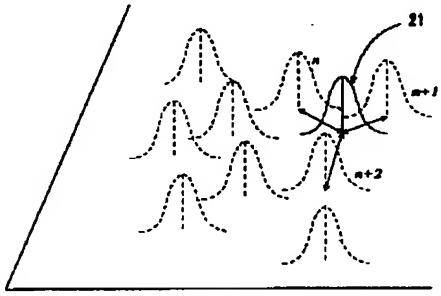
＊【符号の説明】

１～４	HMMの状態
１０	音響解析部
１１	音韻照合部
１２	言語認識部
１３	音声モデル
１４	言語モデル
１５	適応モデル作成部
１６	適応用サンプル・データ
１７	特定話者モデルの集合
２０	多数の話者の特徴ベクトルの分布
２１	適応対象話者の特徴ベクトルの分布

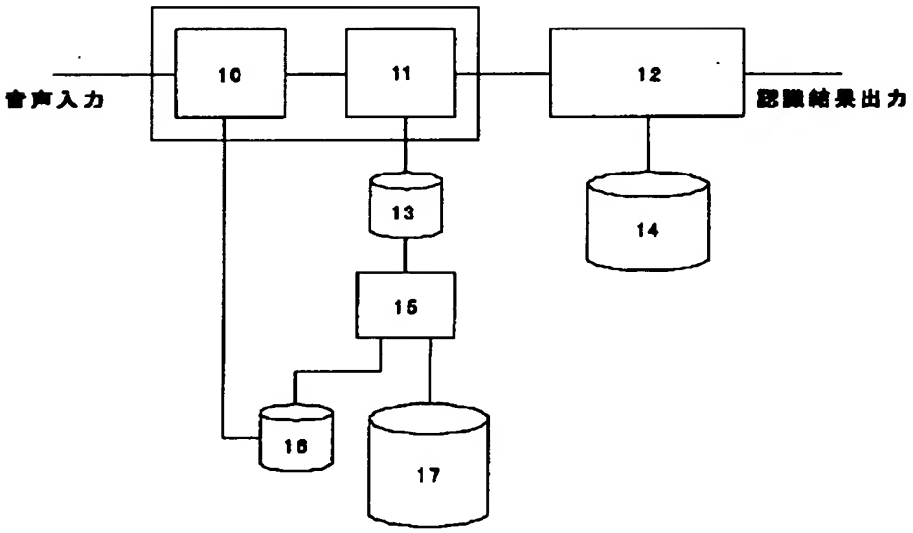
【図１】



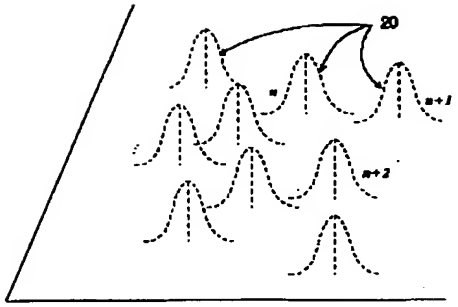
【図４】



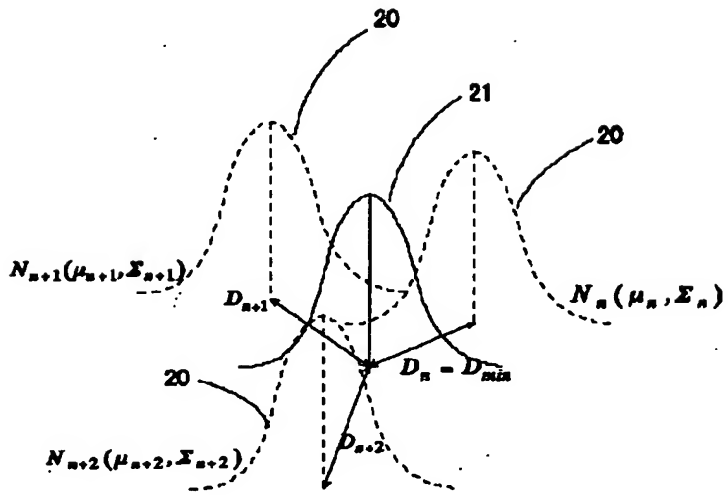
【図２】



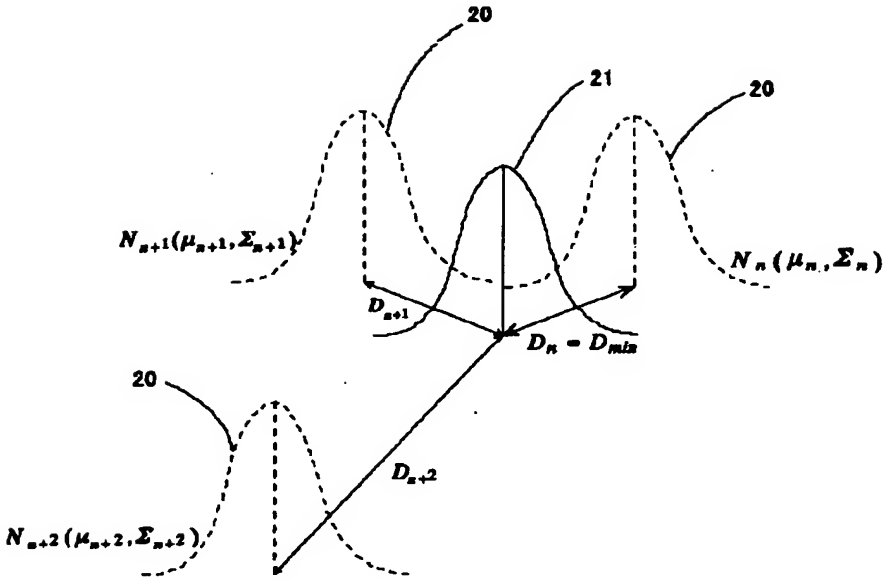
【図３】


20
n
n+1
n+2

【図５】


20
21
20
$N_{n+1}(\mu_{n+1},\Sigma_{n+1})$
$N_n(\mu_n,\Sigma_n)$
D_{n+1}
$D_n = D_{min}$
20
$N_{n+2}(\mu_{n+2},\Sigma_{n+2})$
D_{n+2}

【図６】


20
21
20
$N_{n+1}(\mu_{n+1}, \Sigma_{n+1})$
$N_n(\mu_n, \Sigma_n)$
D_{n+1}
$D_n = D_{min}$
20
D_{n+2}
$N_{n+2}(\mu_{n+2}, \Sigma_{n+2})$